IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

U.S. PATENT APPLICATION

FOR:

METHOD TO IMPROVE PERCEIVED ACCESS SPEED TO DATA NETWORK CONTENT USING A

MULTICAST CHANNEL AND LOCAL CACHE

Inventors:

Rod Walsh
Mika Grundström
Harri Hakulinen

MORGAN & FINNEGAN, L.L.P.
345 PARK AVENUE
NEW YORK, NY 10154-0053

671701 v1

METHOD TO IMPROVE PERCEIVED ACCESS SPEED TO DATA NETWORK CONTENT USING A MULTICAST CHANNEL AND LOCAL CACHE

FIELD OF THE INVENTION

[0001]   This invention relates to systems and methods for sending shared data over a shared access medium.

BACKGROUND OF THE INVENTION

[0002]   Today, many users access content from data networks by using unicast or point-to-point connections. Accessing a data network in this fashion can be slow as a result of delays imposed by the network. Moreover, accessing a data network using point to point connections over a wireless network can be even slower due to delays imposed by both the fixed network and the wireless network.

[0003]   In addition, point-to-point connections over a broadcast network, such as DVB-T (digital video broadcasting – terrestrial), means that the bandwidth must be split between the clients. This is an inefficient use of bandwidth which presents a problem because the DVB-T broadcast network has a limited and shared capacity.

SUMMARY OF THE INVENTION

[0004]   A technical advance is achieved in the art by providing a method to increase a user's perceived access speed to content available from a data network (e.g., internet, extranet, intranet, LAN (local area network), etc.)

[0005]   An exemplary method for increasing perceived access speed to content available from a data network comprises selecting data to be sent to multicast groups based on a predetermined policy and sending the data over the multicast channel. The policy may be, for example, based on client requests or based on information from an agent that monitors web hits from the system clients.

1

[0006]   An exemplary method for increasing perceived access speed to content available from a data network comprises selecting the data to be sent over a shared multicast channel, sending the shared data over the multicast channel, receiving the shared data by a reception agent, filtering the shared data by the reception agent, storing the filtered data in a local cache, and retrieving the cached data for user consumption.

[0007]   In an alternate embodiment, the method involves measuring user demand for data, selecting the data to be sent over a shared multicast channel, sending the shared data over the multicast channel, receiving the shared data by a reception agent, filtering the shared data by the reception agent, storing the filtered data in a local cache, and retrieving the cached data for user consumption.

[0008]   In accordance with another embodiment of the invention, the method involves measuring user demand for data using a web proxy, selecting the data to be sent over a shared multicast channel, sending the shared data over the multicast channel, receiving the shared data by a reception agent, filtering the shared data by the reception agent, storing the filtered data in a local cache, and retrieving the cached data for user consumption.

[0009]   In accordance with another embodiment of the invention, the method involves measuring user demand for data, selecting the data to be sent over a shared multicast channel, sending the shared data over the multicast channel, receiving the shared data by a reception agent, filtering the shared data by the reception agent, storing the filtered data in a local cache, implementing the client with a local web proxy, which will also store data in the local cache, and retrieving the cached data for user consumption.

[0010]   In accordance with yet another embodiment of the invention, the apparatus involves a memory having program code stored therein and a processor connected to said memory for carrying out instructions in accordance with stored program code. Stored program code, when executed by said processor, causes said processor to perform the following steps. Said processor receives a user input request for data from a data network, determines whether said requested data is to be retrieved from the local cache or the data network, and retrieves said requested data for user consumption.

[0011]   Other and further aspects of the present invention will become apparent during the course of the following description and by referring to the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The disclosed inventions will be described with reference to the accompanying drawings, which show exemplary embodiments of the invention and which are incorporated in the specification hereof by reference, wherein:

FIG. 1 depicts an embodiment of the invention where the group data selection is configured by the multicast network administrator.

FIG. 2 depicts an embodiment of the invention where group data is selected according to web hits from users and data network content is retrieved independently of individual user transactions.

FIG. 3 depicts an embodiment of the invention where the individual user transactions are sent as group data.

FIG. 4 depicts a slightly different client implementation using a local web proxy.

FIG. 5 depicts an embodiment of the invention where some clients are active (i.e., client uses the interaction network to send requests) and some clients are passive.

FIG. 6 depicts an exemplary general purpose computer.

DETAILED DESCRIPTION

[0012]   FIG. 1 depicts a system in accordance with one embodiment of the invention with group data selection 110 configured by a multicast network administrator. Group data is data that is selected to be sent over the multicast network 114. The multicast network 114 may employ protocols including DVB-T to transmit the data. As shown in FIG. 1, the multicast

3

system 100 comprises software modules for selecting 110, collecting 108, and sending 112 group data from a data network 104, a database 106 for storing shared content and a multicast network 114. A client terminal 101 for consuming group data may comprise a PC, mobile phone or the like, includes a browser and / or other applications for browsing and consuming content. It also includes a local cache 118 for storing content. In accordance with the present invention, the client terminal 101 also includes a reception agent 116 for receiving and filtering the content. In the embodiment depicted in FIG. 1, an exemplary method for increasing a user's perceived access speed to content available from a data network 104 includes selecting data 110 to be sent over a shared multicast channel 114, collecting the data to be sent 108, optionally sending the collected data to a shared cache 106, sending the data 112 over a multicast network 114, receiving the data by a reception agent 116, filtering the data by the reception agent 116, storing the filtered data in a local cache 118, and retrieving the filtered data from the cache for user consumption.

[0013]   The data network 104 shown in FIG. 1 may be, for example, the internet, an intranet, an extranet, or a LAN. The group data is selected 110 based on a configuration from the multicast network administrator. In one embodiment, the group data may be manually selected and entered into the Multicast system. Once the data is selected, the system would automatically perform the collecting, sending, etc. In another embodiment, the data may be automatically selected using an agent (software) using general policies. For example, the policy (in human readable format) may be:

> send the top 100 web file downloads, the top 100 web sites to a maximum link depth of 20, and the top 101-500 web sites to a maximum link depth of 3. Send only files that are under 1MB in size. Top 1000 chart lists are available from http://www.statistics.com/webtop1000.isam. Update your information every 2 hours.

[0014]   Then, the method of selecting would be performed autonomously by the agent.

[0015]   Once data has been selected, the group data selection module 110 could request that the group data collection module 108 go collect the data. After the data is collected 108, it may be sent directly to the multicast group data sender module 112 or it may be collected in advance of sending over the multicast network 114 by sending the data to the shared cache 106. Once the data is sent over the multicast network 114, the data is received by a reception agent

4

671701 v1

module 116. The reception agent 116 filters the incoming data so that only desirable content for a particular user is stored, and then only up to a certain amount of storage space. The reception agent 116 deals with receiving all data, filtering out the unwanted data, and sending the remainder to the local cache 118. The filtering mechanism may be implemented in a number of ways. In one embodiment, the filtering mechanism can be implemented by using a method of pre-selection specified by the user. For example, content would be advertised ahead of time and a user would select the items they wished to receive. In another embodiment, the filtering mechanism can be implemented with a user configured profile. For example, the user would tell the agent which kind of content he likes (e.g. all web sites with military aircraft and all flight simulator games) and the user would change the profile as he wished. The agent would then automatically filter that kind of content. In another embodiment, the filtering mechanism may be implemented by using a learned user profile. A client agent may actively monitor a user's behavior and construct a profile from that behavior. For example, frequent downloads of 'claudia*.gif' may also prompt reception of 'claudia*.jpg'.

[0016]   The client stores the filtered data in a local cache 118. The local cache 118 may be implemented in a number of ways. In one embodiment, the local cache 118 may be implemented as just a directory structure with web pages and files copied in. In another embodiment, the local cache 118 may be a standards based cache system such as the uHTTP (unidirectional HTTP) cache used by ATVEF (Advanced Television Enhancement Forum). In yet another embodiment, the local cache 118 may use an existing cache system, such as that used by MS Explorer. This allows easy interoperability with the existing application.

[0017]   The filtered data is then retrieved from the local cache 118 for user consumption. The data may be retrieved for use by a client's browser 120 or another client application 124. If the browser application 120 does not find the data in the local cache 118, it could retrieve it directly from the data network 104.

[0018]   FIG. 2 depicts a system similar to FIG. 1. In FIG. 2 the group data is selected 110 according to web hits from users and data network content 102 is retrieved independently of individual user transactions. The multicast system 200 in FIG. 2 differs from that of FIG. 1 in that there is a web hit monitor 202 and an interaction network 204. The interaction network 204 may be a private telephone network, VPN (Virtual Private Network), etc. It may be any bi-

5

directional network. It may be IP (Internet Protocol) over PSTN (Public Switched Telephone Network) or IP over GPRS (General Packet Radio Service) or IP over UMTS (Universal Mobile Telecommunications System) or IP over WLAN (wireless LAN). However, many other options are possible. In this embodiment, the method for increasing a user's perceived access speed to content available from a data network 104 includes measuring user demand for data using a web hit monitor 202, selecting 110 data to be sent over a shared multicast channel 114 based upon said user demand, collecting the data to be sent 108, optionally sending the collected data to a shared cache 106; sending the data 112 over a multicast network 114, receiving the data by a reception agent 116, filtering the data by the reception agent 116, storing the filtered data in a local cache 118, and retrieving the filtered data from the cache 118 for user consumption.

[0019]   In the embodiment of FIG. 2, user demand is measured by using a web hit monitor 202. The monitoring can be performed in a number of ways. In one embodiment, the web hit monitor may include a proxy. The data passes through the proxy and the requests for content are made by the proxy on behalf of the client. A visible proxy may be accessed by configuring the client software (e.g. setting up the proxy on a browser or client-local proxy). In this way, the client ensures that all its requests (e.g. in IP packets) are routed via that proxy. A transparent proxy is not known by a client so another mechanism, such as IP tunneling or PPP (Point-to-Point Protocol) over a dial-up connection, must be used to ensure that client requests are 'routed' via this proxy. The term proxy is used generically. An implementation may use several functional proxies to perform the role of the 'logical proxy' described here (e.g., separate FTP (File Transfer Protocol) and HTTP (HyperText Transport Protocol) proxies may be used in the actual implementation.)

[0020]   In another embodiment, web hits may be monitored by using a "snooper." A snooper can 'see' data on a connection as it passes, e.g. HTTP/TCP(Transmission Control Protocol)/IP packets on an Ethernet subnet. Unlike a proxy, the packets do not 'pass through' the Snooper. Like a transparent proxy, some method of routing the packets via a 'snooped subnet' is necessary, e.g. terminating the server side of an L2TP (Layer 2 Tunneling Protocol) tunnel (between the network and client) in the same domain as the Snooper.

[0021]   In another embodiment, web hits may be monitored by using a Network Management Agent. Network management software and protocols provide a means to locally

6

and remotely read data about data traffic. For example, routers may monitor all data that passes through them provide information relating to this data to a network management agent using a communication protocol such as SNMP (Simple Network Management Protocol). This network management agent may retrieve traffic reports in real-time or periodically from the router or some other network element (e.g. a VPN server or HTTP proxy) and detect client web requests from the reports.

[0022] As shown in FIG. 2, the data to be sent over the multicast channel 114 is selected by the group data selection module 110 based upon the information gathered by the web hit monitor module 202. For example, after a predetermined number of clients have requested the same web page then the page is selected to be sent over the multicast channel 114. Another example is that when a number of hits to a "sponsored page" are made, a promotional file (e.g. game) is also sent. Moreover, the thresholds may be location related, e.g. it must be two people in a certain (DVB-T) broadcast cell or on a certain IP subnet.

[0023] In another embodiment, the data may be selected by using the clients' previous web hits as a guide. For example, a hit to <u>www.nokia.com/index.html</u> would lead the group data selection module to predict that all the links (and images) on that page should also be sent. Another example is where the historical behavior is known (i.e., recorded) and the prediction is based on the 'normal' choice of subsequent web hits. Another example would be where a 'sponsor' policy is available and sponsored data is treated as preferable (e.g., send the linked page as well as the requested page either 'if it belongs to a sponsor and there is a 10% chance of being followed' or 'if it does not belong to a sponsor and there is a 50% chance of it being followed.') Moreover, some of these examples can be combined to create equally valid uses of the present invention.

[0024] FIG. 3 depicts a system similar to FIG. 2. In FIG. 3 the individual user transactions are sent as group data. The multicast system 300 in FIG. 3 differs from that of FIG. 2 in that a web proxy 302 replaces the following three elements: the web hit monitor module 202, the group data selection module 110, and the group data collection module 108. In this embodiment, an exemplary method for increasing a user's perceived access speed to content available from a data network 104 includes selecting data to be sent over a shared multicast channel 114 by using a unicast to multicast proxy 302, sending the data 112 over a multicast

7

network 114, receiving the data by a reception agent 116, filtering the data by the reception agent 116, storing the filtered data in a local cache 118, and retrieving the filtered data from the cache 118 for user consumption.

[0025]   By employing a unicast to multicast proxy 302, the items to be included in group data may be selected directly from normal client requests for unicast delivery (e.g., HTTP / TCP request) and the items may be taken directly from the unicast delivery stream. The unicast to multicast proxy 302 can be configured in a number of ways. In one embodiment, the unicast would only be converted into multicast after a threshold number of hits. In another embodiment, all client requests for unicast may be sent by multicast. In yet another embodiment, the unicast would only be converted into multicast for selected sites identified by a URL (Uniform Resource Locator) stub or subnet address. In still another embodiment, the unicast would only be converted into multicast if a predetermined period of time has elapsed since the content was last sent (e.g., in the last 12 hours.)

[0026]   FIG. 4 depicts a client implementation of the present invention using a local proxy 402. The client implementation of FIG. 4 differs from the client implementation of FIG. 3 in that a local proxy 402 is used in FIG. 4. In this implementation, the group data sent by the multicast system is filtered by the reception agent 116, the filtered data is sent to the local cache 118, and the filtered data from the cache 118 is available for use by a browser or other application. Also, the data network 104 content is retrieved by the browser 120 or other applications 124 from a local cache 118 or the data network via a local proxy 402. The local proxy 402 may interact with the reception agent 116. In one embodiment, the local proxy 402 and the reception agent 116 may interact through an API (Application Programming Interface). A "nice" API may have the local proxy 402 ask the reception agent 116 if the client requested data is coming. In addition, an API may also ask about data that the reception agent predicts (e.g., Squid has some of this predictive functionality.) In another embodiment, the local proxy 402 and the reception agent 116 may be part of the same code and just work on internal function calls (as most WebCrawlers do.) In yet another embodiment, the local proxy 402 and the reception agent 116 may interact using any interprocess communication (e.g., Java-RMI (Remote Method Invocation), ONC (Open Network Computing)-RPC (Remote Procedure Call), SOAP (Simple Object Access Protocol) , CORBA (Common Object Request Broker Architecture), DCOM (Distributed Component Object Model), TCP sockets, etc.)

8

671701 v1

[0027]   The installation of a local proxy 402 on the client system would solve a number of problems. For example, the user would not need to do any special configuration because application requests for network data would pass through the proxy 402. Thus, for example, if www.demogames.com is browsed, the proxy 402 can choose whether to make an Internet request or whether to retrieve the data from the cache 118 (if available). In addition, there can be special proxy/receiver messaging and metadata which enables special features. A useful special feature would be to indicate when the cached content is out-of-date (i.e., the Internet content has changed on the Internet Server). This may be implemented as a "data will expire" time stamp or an actual announcement from the multicast system. Moreover, a proxy 402 is also a way to allow a site (e.g., either a FTP or HTTP site) to be partially or completely mirrored. Mirroring is a common technique on the internet.

[0028]   FIG. 5 depicts an embodiment of the invention where some clients are active (i.e., client uses the interaction network to send requests) and some clients are passive. According to the present invention, some or all clients may be in either an active 506, 508 or a passive 502, 504 state. A client is active 506, 508 when it uses the interaction network 204 to send requests. A client is passive 502, 504 when it is able to receive data from the Multicast Network 114 but it is not using the interaction network 204 for this service. Note that the state (active or passive) of a client may change with time.

[0029]   The software, modules, components and other code and/or software elements described above could be written, for example, using an object-oriented language known in that art such as Objective-C, Java, or C#. Phrases such as "module" as used herein, may refer, for example, to program code written as a class using an object-orientated programming language and instantiated into an object using techniques known in the art of object-orientated programming.

[0030]   The modules or other software could run, for example, on computers including personal computers or workstations such as Power Macintosh G4s or Dell Dimensions running operating systems such as Apple OS X, Microsoft Windows XP, or Linux, perhaps further including support for Java. The modules or other software could also run, for example, on PDAs (Personal Digital Assistant), cellular telephones, DVB-T receivers, or the like running an operating system such as Microsoft Windows CE or Symbian EPOC, perhaps with support for

9

Java. Speaking, more generally, the modules or other software could run on a general purpose computer.

[0031]   The phrases "general purpose computer," "computer," and the like, as used herein, refer but are not limited to an engineering workstation, PC, Macintosh, PDA, web-enabled cellular phone and the like running an operating system such as OS X, Linux, Windows CE, Windows XP, Symbian EPOC, or the like. The phrases "general purpose computer," "computer," and the like also refer, but are not limited to, one or more processors operatively connected to one or more memory or storage units, wherein the memory or storage may contain data, algorithms, and/or program code, and the processor or processors may execute the program code and/or manipulate the program code, data, and/or algorithms. Accordingly, exemplary computer 600 as shown in Fig. 6 includes system bus 650 which operatively connects two processors 651 and 652, random access memory (RAM) 653, read-only memory (ROM) 655, input output (I/O) interfaces 657 and 658, storage interface 659, and display interface 661. Storage interface 659 in turn connects to mass storage 663. Each of I/O interfaces 657 and 658 may be an Ethernet, IEEE 1394, IEEE 802.11, or other interface such as is known in the art. Mass storage 663 may be a hard drive, optical disk, or the like. Processors 651 and 652 may each be a commonly known processor such as an IBM or Motorola PowerPC, or an Intel Pentium.

[0032]   Computer 600 as shown in this example also includes an LCD display unit 601, a keyboard 602 and a mouse 603. In alternate embodiments, keyboard 602 and/or mouse 603 might be replaced with a pen interface. Computer 600 may additionally include or be attached to card readers, DVD drives, or floppy disk drives whereby media containing program code may be inserted for the purpose of loading the code onto the computer. In accordance with the present invention, computer 600 may be programmed using a language such as Java, Objective C, C, C#, or C++ according to methods known in the art to perform those operations described above.

[0033]   Accordingly, the above described user terminal could be, for example, a portable device comprising a StrongARM processor, an integrated touch-sensitive color screen with the ability to receive DVB-T broadcasts and, in some embodiments, the ability to send and receive GSM, PCS, or other cellular transmissions. The device could use an operating system such as Microsoft Windows CE or Symbian EPOC, perhaps with support for Java.

10

671701 v1

[0034]   The many features and advantages of the present invention are apparent from the detailed specification, and thus, it is intended by the appended claims to cover all such features and advantages of the invention which fall within the true spirit and scope of the invention.

[0035]   Furthermore, since numerous modifications and variations will readily occur to those skilled in the art, it is not desired that the present invention be limited to the exact construction and operation illustrated and described herein, and accordingly, all suitable modifications and equivalents which may be resorted to are intended to fall within the scope of the claims.